

TECHNICAL NOTES

The Occupational Employment Statistics (OES) survey is an annual mail survey of occupational employment and wage rates for wage and salary workers in nonfarm establishments, by industry. Approximately 400,000 establishments are sampled for the survey each year; over 3 years, approximately 1.2 million establishments are contacted. The reference period for each year's survey is the fourth quarter of that year. Although estimates can be made from a single year of data, the OES survey has been designed to produce estimates using a full 3 years of data. The sample allows the production of estimates at detailed levels of geography, industry, and occupation. (See Estimation, below.)

Extensive portions of the material in these technical notes have been excerpted or reproduced verbatim from "Appendix B. Survey Methods and Reliability of the 2001 Occupational Employment Statistics Estimates" of Bulletin 2559, *Occupational Employment and Wages, 2001* (June 2003; available online at http://bls.gov/oes/2001/appendix_b.pdf), of the U.S. Department of Labor, Bureau of Labor Statistics (BLS). Readers are encouraged to consult that appendix for more complete explanations.

OCCUPATIONAL CLASSIFICATION

The 1999 OES survey was the first to incorporate the Standard Occupational Classification System (SOC), a revised occupational classification system of the Office of Management and Budget (OMB). The SOC is the occupational classification system required by OMB for use by all federal agencies. The OES survey uses 22 major occupational groups from the SOC to categorize workers in one of almost 770 detailed occupations. The 2001 OES survey wage estimates were developed from combined 1999, 2000, and 2001 data obtained from an initial sample of 1,208,542 establishments. Occupational employment estimates are based only on data collected in the 2001 survey.

The major groups of the SOC are as follows:

- Management occupations
- Business and financial operations occupations
- Computer and mathematical occupations
- Architecture and engineering occupations
- Life, physical, and social science occupations
- Community and social services occupations
- Legal occupations
- Education, training, and library occupations

- Arts, design, entertainment, sports, and media occupations
- Healthcare practitioners and technical occupations
- Healthcare support occupations
- Protective service occupations
- Food preparation and serving related occupations
- Building and grounds cleaning and maintenance occupations
- Personal care and service occupations
- Sales and related occupations
- Office and administrative support occupations
- Farming, fishing, and forestry occupations
- Construction and extraction occupations
- Installation, maintenance, and repair occupations
- Production occupations
- Transportation and material moving occupations
- Military-specific occupations (not surveyed as part of OES).

DEFINITIONS

Employment. Employment is defined as the number of workers who can be classified as full-time or part-time employees, including workers on paid vacations or other types of leave; workers on unpaid short-term absences; salaried officers, executives, and staff members of incorporated firms; employees temporarily assigned to other units; and employees for whom the reporting unit is their permanent duty station, regardless of whether that unit prepares their paycheck. Among those excluded from coverage are most proprietors (owners and partners of unincorporated firms), self-employed workers, and unpaid family workers. Employees are reported in the occupation in which they are working, rather than the occupation for which they were trained.

In this report, employment represents the estimate of total wage and salary employment in an occupation. To reduce paperwork and respondent burden, no OES survey form contains every SOC occupation. Instead, the survey form sent to an establishment contains 50 to 225 SOC occupations selected on the basis of the industry classification and size class of the sampled establishment. Thus, data for specific occupations are collected primarily from establishments within industries that are the predominant employers of labor in those occupations. Occupations not listed can be added to the survey form.

Establishment. An establishment is an economic unit that produces goods or services. It generally is found at a

single physical location and is engaged predominantly in one type of economic activity. Where a single physical location encompasses two or more distinct activities, these are treated as separate establishments if separate payroll records are available and certain other criteria are met.

Standard Industrial Classification (SIC). The industrial classification system used in this survey is described in the *Standard Industrial Classification Manual: 1987* (Office of Management and Budget: Washington, DC), which classifies reporting establishments into industries on the basis of major product or activity. The OES program produces estimates by both two-digit and three-digit SIC codes, estimates across all industries, and estimates of total national employment.

Wages. Wages for the OES survey are straight-time, gross pay, exclusive of premium pay. Base rate, cost-of-living allowances, guaranteed pay, hazardous-duty pay, incentive pay including commissions and production bonuses, tips, location differential, length-of-service allowances, and on-call pay are included. Excluded are attendance bonuses, back pay, jury duty pay, overtime pay, severance pay, shift differentials, nonproduction bonuses, tuition reimbursements, meal and lodging allowances, merchandise discounts, profit-sharing distributions, relocation allowances, and stock bonuses.

The OES survey collects wage data in 12 intervals. Employers report the number of employees in an occupation by wage interval. The wage intervals used for the 2001 survey are as follows:

Interval	Wages (dollars)	
	Hourly	Annual
A	under 6.75	under 14,040
B	6.75–8.49	14,040–17,679
C	8.50–10.74	17,680–22,359
D	10.75–13.49	22,360–28,079
E	13.50–16.99	28,080–35,359
F	17.00–21.49	35,360–44,719
G	21.50–27.24	44,720–56,679
H	27.25–34.49	56,680–71,759
I	34.50–43.74	71,760–90,999
J	43.75–55.49	91,000–115,439
K	55.50–69.99	115,440–145,599
L	70.00 and over	145,600 and over

Mean wage. The mean wage is the estimated total wages for an occupation divided by its weighted survey employment. A mean hourly wage value is calculated for each wage interval, A through K, based on occupational wage data collected by the BLS Office of Compensation and Working Conditions. The mean wage value for the upper open-ended wage interval L (\$70.00 and over) is its lower bound (Winsorized mean). These interval mean wage values are then attributed to all workers reported in the interval. For each occupation, total weighted wages in each interval are summed across all intervals and divided by the occupation's weighted survey employment.

Median wage. The median wage is the estimated 50th percentile of the distribution of wages: 50 percent of workers in an occupation earn wages below, and 50 percent earn wages above, the median wage. The wage interval containing the median wage is located using a cumulative frequency count of employment across wage intervals. The median wage rate is then estimated using a linear interpolation procedure.

Annual wage. Annual wage estimates are calculated by multiplying the mean hourly wage by 2,080 hours (52 weeks per year multiplied by 40 hours per week). Employees paid at an hourly rate by their employers may work less than or more than 40 hours per week. Thus, the annual wage estimates may not represent the actual annual pay received by employees. For a small number of occupations in this report only an annual wage figure is provided. The workers in these occupations are generally paid on an annual basis, and their annual wage has been directly calculated from the reported survey data.

Producing estimates using 3 years of sample data provides additional occupational detail and reduces sampling error (particularly for small geographic areas and occupations). However, this procedure also has quality limitations because it requires the adjustment of data from earlier years to the current reference period—a procedure referred to as “wage updating.” The OES program uses the over-the-year fourth-quarter wage changes from the BLS Employment Cost Index (ECI) to adjust prior-year survey data before combining them with the current-year data. The wage updating procedure assumes

that each occupation's wage, as measured in the earlier years, moves according to the average movement of its occupational division and that there are no major geographic or detailed occupational differences—and this may not be the case. BLS has conducted research over the past several years on the accuracy of the ECI wage-updating method compared with other modeling approaches. Current research results support the continued use of the ECI wage-updating methodology.

SCOPE OF SURVEY

The survey covers establishments in SIC codes 07, 10 through 42, 44 through 87, 89, and state and local governments. In addition, data for the U.S. Postal Service and for the federal government are universe (total) counts obtained from the U.S. Office of Personnel Management (OPM). Occupational employment and wage estimates at the national level were produced by BLS using employment and wage data from the 50 U.S. states and the District of Columbia. Guam, Puerto Rico, and the U.S. Virgin Islands were surveyed; however, data from these territories are not included in the production of national estimates.

For the OES survey, employers are requested to provide occupational data for a particular reference date. The reference date for any particular establishment in the survey is dependent on its SIC code. The reference date for the 2001 survey was the pay period that included October 12, November 12, or December 12 of 2001, depending on SIC code. The pay period including the 12th day of the reference month is standard for federal agencies collecting employment data.

METHOD OF COLLECTION

Survey questionnaires (schedules) were initially mailed out to almost all sampled establishments; personal visits were made to some of the larger establishments.

Two additional mailings were sent to nonresponding establishments at approximately 3-week intervals. Telephone or personal-visit follow-ups were made for those nonresponding establishments considered critical to the survey because of their size.

SAMPLING PROCEDURES

The OES survey is based on a probability sample and is designed to represent the universe of establishments it covers. The survey is conducted over a 3-year

cycle. Each year, one-third of the sample units are included in the survey. To the extent possible, units selected in 1 year are not included in the sample the following 2 years.

Establishments in eligible two- and three-digit SIC codes that reported to a state employment security agency for unemployment insurance purposes constitute the sampling frame for this survey. Virtually all businesses are required to file such a report with the state in which they are located. Each quarter, BLS combines the lists from all states into a single file called the Longitudinal Database (LDB), a compilation of state unemployment insurance reports. For the 1999 survey the sampling frame was the LDB file from the second quarter of 1998, for the 2000 survey it was the LDB file from the second quarter of 1999, and for the 2001 survey it was the LDB file from the fourth quarter of 2000. The sampling frame was supplemented with a list supplying establishment information on railroads (SIC 401). OPM provided data representing federal government employment and wages, obtained from an annual census of federal government establishments, at the end of the survey process.

Within each state, establishments in the universe were stratified by Metropolitan Statistical Area (MSA), three-digit SIC code, and size of firm. An establishment's size class is determined by its employment as reported on the sampling frame. Establishments in smaller size classes were selected based on a probability sample. Establishments in larger size classes are sampled with virtual certainty during the 3-year cycle of the survey. The targeted sample size of 1.2 million establishments per 3-year cycle was allocated in a manner that equalized the expected relative standard error of the typical occupational employment within the cell for each MSA and three-digit SIC. Within each of these cells, the sample was allocated across size classes in a manner that minimized the variance of the average typical occupational employment estimate.

RESPONSE

Of the 369,694 eligible units from the 1999 sample, usable responses were obtained from 286,903, producing a response rate of 77.6 percent based on units. Of the 375,387 eligible units from the 2000 sample, usable responses were obtained from 293,450, producing a response rate of 78.2 percent based on units. Of the 366,760 eligible units from the 2001 sample, usable responses were obtained from 286,726, producing a response rate of 78.2 percent based on units.

ESTIMATION

Combining data across years was challenging because of the 1999 transition to a new SOC-based OES occupational coding system. Although most of the former OES occupations can be crosswalked to a counterpart in the new system, many of the relations between the two coding systems are not one-to-one. Many former OES occupations are crosswalked to residual occupations, meaning that occupation is no longer surveyed as a detailed occupation. For more information about the SOC, please see the discussion of the SOC at the BLS Web site (<http://www.bls.gov/soc>).

SAMPLE WEIGHTS

Each sampled establishment was assigned an original sampling weight, the reciprocal of the establishment's probability of selection (i.e., its design weight) within its sampled year.

Weights were modified for each in-scope establishment in a cell by dividing the establishment's design weight by a factor indicating the number of years for which sample units were selected from that sampling cell. This weight was used in the calculation of the 2001 estimates based on combining data from the 1999, 2000, and 2001 surveys.

NONRESPONSE

Nonresponding establishments are accounted for in the OES survey by a two-step imputation process. First, the staffing pattern is imputed using a "hot-deck," "nearest-neighbor" imputation method. Hot-deck procedures use data from the current period to impute for missing data (from the current period). The nearest-neighbor method searches the responding establishments within a defined cell and finds the one that most closely matches the nonresponding establishment for key classification values (such as area, SIC, size class). The staffing pattern (employment distribution), of the responding establishment is used as the staffing pattern of the nonresponding establishment.

COMBINING AND BENCHMARKING MULTIYEAR DATA

Whenever possible, data from the 1999, 2000, and 2001 surveys were combined. The remaining occupational wage estimates and all of the employment estimates were produced using only 2001 data. Each year's sample was weighted to represent the sample as it appeared at the time the sample was selected. In order to

combine the data, each unit's weight was modified to have the aggregate sample represent the universe. This was done by dividing each unit's weight by the number of years for which sample units were selected for that stratum.

ESTIMATED EMPLOYMENT

A ratio estimator was used to develop estimates of occupational employment. The auxiliary variable was the population value of total employment obtained from the refined unemployment insurance files for the 2001 reference month. Within each MSA, the estimated employment for an occupation at the reported three-digit SIC level was calculated by multiplying the weighted employment by its ratio factor. The estimated employment for an occupation at the all-industry level was obtained by summing the occupational employment estimates across all industries within an MSA reporting that occupation. The employment and wage data for federal government workers in each occupation were added to the survey-derived data.

VARIANCE OF ESTIMATES

Estimates of sampling error are calculated to allow the users to determine if occupational employment estimates are reliable enough for their needs. Only a probability-based sample can be used to calculate estimates of sampling error from the sample itself.

The formula used to estimate occupational employment variances (a common measure of sampling error) is based on the survey's sample design and method of estimation. The OES survey used a subsample replication technique called the jackknife random group to estimate variances of occupational employment. In this technique, each sampled establishment is assigned to one of G random groups. Using the data in these groups, G subsamples are formed from the parent sample. Next, G estimates of total employment for an occupation P are calculated, one employment estimate per subsample. The variability of these G employment estimates is then calculated. This variability is the BLS variance estimate of the employment estimate for occupation P .

DISCREPANCIES BETWEEN EMPLOYMENT ESTIMATES AND WAGE ESTIMATES

Users consulting both occupational employment estimate tables and wage estimate tables may notice apparent discrepancies between two tables in the treatment of identical variables. For instance, wage estimates may

be displayed for certain occupations for which no employment estimates are reported, or employment or wage data may be displayed at the two-digit SIC level but not for the component three-digit SIC industries that together constitute the displayed two-digit industry. The two principal reasons for apparent discrepancies are (1) that BLS-applied suppression rules differ for employment estimates and for wage estimates, and (2) data at the three-digit SIC level may have to be suppressed to assure that individual establishments cannot be identified.

RELIABILITY OF THE ESTIMATES

Estimates developed from a sample may differ from the results of a census. Two types of error, sampling and nonsampling, can occur in estimates calculated from a sample. Sampling error occurs because observations are based on a sample, not on the entire population. Nonsampling error occurs because of response and operational errors in the survey. Unlike sampling error, this form of error can also occur in a census.

SAMPLING ERROR

The particular sample used in this survey is one of many possible samples of the same size that could have been selected using the same sample design. Estimates derived from different samples tend to differ from one another. The variance of a survey estimate is a measure of the variation among the estimates from all possible samples. The standard error of a survey estimate is the square root of its variance; the relative standard error is the ratio of the standard error to the estimate itself.

By using the sample estimate and its standard error, the user can construct an interval estimate with a prescribed level of confidence that the interval will include the mean value of the estimate from all possible samples.

For example, suppose that an estimated occupational employment total is 5,000 and has an associated relative standard error of 2.0 percent. Based on these data, the standard error of the estimate is 100 (2 percent of 5,000). A 68 percent confidence interval for the employment estimate is $5,000 \pm 100$, or from 4,900 to 5,100. Approximately 68 percent of the intervals constructed in this manner will include the mean of all possible employment estimates as computed from all possible samples. A 95 percent confidence interval for the employment estimate is $5,000 \pm 196$, or from 4,804 to 5,196. Approximately 95 percent of the intervals constructed in this manner will include the mean of all possible employ-

ment estimates as computed from all possible samples. Estimates of sampling errors for occupational employment estimates are available for most estimates.

NONSAMPLING ERROR

Nonsampling error is attributable to such causes as an inability to obtain information for all establishments in the sample; differences in respondents' interpretation of the survey question; respondents' inability or unwillingness to provide correct information; errors made in recording, coding, or processing the data; and errors made in imputing values for missing data. Explicit measures of the effects of nonsampling error are not available. The relative standard error indicates the magnitude of the sampling error; it does not measure nonsampling error, which includes biases in the data. Particular care should be exercised in the interpretation of small estimates or of small differences between estimates when the sampling error is relatively large or the magnitude of the bias is unknown.

Several edit and quality-control procedures were used to reduce nonsampling error. For example, completed survey questionnaires were checked for data consistency, follow-up mailings were sent to nonresponding establishments to improve the survey response rate, and response analysis studies were conducted to assess respondents' comprehension of the questionnaire. Additional quality control procedures used in the OES survey are described below in "Quality Control Measures."

RELATIVE STANDARD ERROR NOT DISPLAYED

Mean hourly wages are calculated from the mean values of the lower 11 of 12 wage intervals using data from the BLS National Compensation Survey (see Definitions, above). Because of space restrictions, relative standard errors are not displayed for estimates of mean hourly wages and mean annual wages for scientists, engineers, and technicians in tables 13–20. Relative standard errors for mean hourly wages were calculated and are available on request. Relative standard errors were not calculated for mean annual wages because the estimates for mean annual wages were calculated directly by multiplying mean hourly wages by 2,080 hours, which for this survey represents full-time employment.

All employment estimates for employees not allocated to a specific SIC (tables 1–4 and table 10) are residually determined by subtracting the subtotal of

estimates allocated by industry from the estimate of total filled positions. Because these values are calculated rather than estimated, no relative standard error of the estimate is shown for them in table 10. Relative standard errors of the employment estimates are displayed for occupational subclassifications in tables 5–10 but not for the occupational totals. Relative standard errors of these estimates are not available because the occupational totals are simple arithmetic sums of the occupational subclassification estimates.

QUALITY CONTROL MEASURES

The OES survey is a cooperative program and has limited personnel resources. Nonetheless, the program must accommodate state-specific publication needs; standardize survey procedures across all 50 U.S. states, the District of Columbia, and the U.S. territories; and produce quality estimates. Controlling sources of nonsampling error in this decentralized environment can be difficult. In addition, edit and validation checks are distributed across eight regional offices, which can lead to procedural differences between the regions. Two important quality control measures used by the OES survey are the Survey Processing and Management (SPAM) System and the Estimates Delivery System (EDS). Both systems were developed to provide a consistent and au-

tomated framework for survey processing and to reduce the workload at the state, regional, and national levels.

By standardizing data processing activities, such as refining mailing addresses, addressing envelopes and mailers, editing and updating questionnaires, producing management reports, and calculating employment estimates, the SPAM system and the EDS have consequently standardized survey methodology. This has reduced the number of errors on the data files as well as the time needed to review them.

Other quality control measures implemented in the OES survey include

- Follow-up of solicitations of nonrespondents (especially critical nonrespondents),
- Review of schedules to verify the accuracy and reasonableness of the reported data,
- Adjustments of atypical reporting units on the data file,
- Validation of the benchmark employment figures and of the benchmark factors,
- Validation of the analytical tables of estimates (at the two- and three-digit SIC levels), and
- Use of bar codes to reduce keypunch errors.